

University of Groningen

Compositional profile of α/β -hydrolase fold proteins in mangrove soil metagenomes

Jiménez Avella, Diego; Dini Andreote, Francisco; Ottoni, Júlia Ronzella; de Oliveira, Valéria Maia; van Elsas, Jan Dirk; Andreote, Fernando Dini

Published in:
Microbial Biotechnology

DOI:
[10.1111/1751-7915.12157](https://doi.org/10.1111/1751-7915.12157)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Jiménez Avella, D., Dini Andreote, F., Ottoni, J. R., de Oliveira, V. M., van Elsas, J. D., & Andreote, F. D. (2015). Compositional profile of α/β -hydrolase fold proteins in mangrove soil metagenomes: Prevalence of epoxide hydrolases and haloalkane dehalogenases in oil-contaminated sites. *Microbial Biotechnology*, 8(3), 604-613. <https://doi.org/10.1111/1751-7915.12157>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Brief report

Compositional profile of α/β -hydrolase fold proteins in mangrove soil metagenomes: prevalence of epoxide hydrolases and haloalkane dehalogenases in oil-contaminated sites

Diego Javier Jiménez,^{1*} Francisco Dini-Andreote,¹ Júlia Ronzella Ottoni,² Valéria Maia de Oliveira,² Jan Dirk van Elsas¹ and Fernando Dini Andreote³

¹Department of Microbial Ecology, Centre for Ecological and Evolutionary Studies, University of Groningen, Groningen 9747AG, The Netherlands.

²Microbial Resources Division, Research Center for Chemistry, Biology and Agriculture, University of Campinas, Campinas, SP 6171, Brazil.

³Department of Soil Science, 'Luiz de Queiroz' College of Agriculture, University of São Paulo, Piracicaba, SP, CEP 13418-900, Brazil.

Summary

The occurrence of genes encoding biotechnologically relevant α/β -hydrolases in mangrove soil microbial communities was assessed using data obtained by whole-metagenome sequencing of four mangroves areas, denoted BrMgv01 to BrMgv04, in São Paulo, Brazil. The sequences (215 Mb in total) were filtered based on local amino acid alignments against the Lipase Engineering Database. In total, 5923 unassembled sequences were affiliated with 30 different α/β -hydrolase fold superfamilies. The most abundant predicted proteins encompassed cytosolic hydrolases (abH08; ~ 23%), microsomal hydrolases (abH09; ~ 12%) and *Moraxella* lipase-like proteins (abH04 and abH01; < 5%). Detailed analysis of the genes predicted to encode proteins of the abH08 superfamily revealed a high proportion related to epoxide hydrolases and haloalkane dehalogenases in polluted mangroves BrMgv01-02-03. This suggested selection and putative involvement in local

degradation/detoxification of the pollutants. Seven sequences that were annotated as genes for putative epoxide hydrolases and five for putative haloalkane dehalogenases were found in a fosmid library generated from BrMgv02 DNA. The latter enzymes were predicted to belong to *Actinobacteria*, *Deinococcus-Thermus*, *Planctomycetes* and *Proteobacteria*. Our integrated approach thus identified 12 genes (complete and/or partial) that may encode hitherto undescribed enzymes. The low amino acid identity (< 60%) with already-described genes opens perspectives for both production in an expression host and genetic screening of metagenomes.

Introduction

Mangroves harbour diverse microbial communities which play critical roles in the functioning and maintenance of these sensitive and complex systems (Kathiresan and Bingham, 2001; Sahoo and Dhal, 2009). Given the peculiar factors that drive these systems (salt, anaerobic/aerobic shifts), mangroves offer rich sources of genes for new biotechnological products/enzymes, such as lipases (Couto *et al.*, 2010), cellulases (Thompson *et al.*, 2013) and laccases (Ye *et al.*, 2010). Mangrove soils have already been explored for microbial diversity using a wide range of culture-dependent and culture-independent methods (Dias *et al.*, 2009; dos Santos *et al.*, 2011). In particular, the modern metagenomics-based tools, i.e. high-throughput sequencing of environmental DNA followed by a directed search for target genes, allow a ready access to the metabolic potential of mangrove microbial communities (Andreote *et al.*, 2012). Moreover, genes for important enzymes can be further custom-synthesized and codon-optimized, after which heterologous expression may be achievable in a suitable host, an approach that has been coined 'synthetic metagenomics' (Chistoserdova, 2010). To access whole operons, metagenomic libraries can be constructed in large-insert vectors and screened either by functional or genetic approaches. Importantly, functional screening analysis

Received 11 April, 2014; revised 22 July, 2014; accepted 24 July, 2014. *For correspondence. E-mail d.j.jimenez.avella@rug.nl; Tel. (+31) 503632191; Fax (+31) 503632348.

Microbial Biotechnology (2015) 8(3), 604–613
doi:10.1111/1751-7915.12157

Funding Information This work was financially supported by the São Paulo Research Foundation (FAPESP, No. 2004/13910-6, 2010/51981-3, 2011/50809-5 and 2012/06245-2).

does not depend on prior sequence information to detect the target proteins, provided these become available and active in the novel host. Recently, it was shown that particular sequence/activity incoherencies in databases can be solved using expression detection (Fernández-Arrojo *et al.*, 2010; Jiménez *et al.*, 2012a). However, proper testing requires that the expression conditions in the heterologous host are adequate (Ekkers *et al.*, 2012).

Current classification of metagenomic sequencing data relies strongly on local alignments (e.g. using BLAST) against public databases (e.g. NCBI, SEED and KEGG) (Montaña *et al.*, 2012). However, completely novel biotechnologically relevant proteins cannot be easily discovered using such approach. Metagenome datasets can be assessed by the use of more specific databases, such as for example the CAZy (Carbohydrate-Active Enzyme) (Cantarel *et al.*, 2008), PeroxiBase (Fawal *et al.*, 2013), Lipase/Laccase Engineering (Fischer and Pleiss, 2003; Sirim *et al.*, 2011), ESTHER (Lenfant *et al.*, 2013), 3DM (Kourist *et al.*, 2010), Epoxide Hydrolases and/or Haloalkane Dehalogenases (Barth *et al.*, 2004). Alternatively, catalytic or structurally conserved domains can be detected using hidden Markov models (HMM). However, most HMM are designed based on protein sequences retrieved from databases and so the true novelty is still questionable.

The α/β -hydrolase fold enzymes, present in the Lipase Engineering Database (LED), constitute a protein family with diverse catalytic and non-catalytic functions. The α/β -hydrolase fold proteins consist of eight β -strands connected by α -helices. These enzymes are characterized by a common catalytic triad formed by a catalytic nucleophile (serine, aspartate or cysteine), a histidine and an acidic residue (aspartate or glutamate). These residues occur on conserved locations in loops and the α/β -hydrolase fold brings them together to form the active site (Lenfant *et al.*, 2013). These proteins encompass several key enzymes for biocatalytic applications, e.g. lipases, esterases, epoxide hydrolases (EHs), C–C breaking enzymes, dehalogenases and hydroxynitrile lyases (Holmquist, 2000).

Interestingly, EHs (Enzyme Commission number-EC 3.3.2.9) from microbial sources have been recently recognized as a versatile group of enzymes that are important for the synthesis of enantiopure oxides and vicinal diols (intermediates in the organic synthesis of chiral pharmaceutical compounds, drugs and agrochemicals) (Lee and Shuler, 2007; Choi, 2009; Sareen and Kumar, 2011). Such EHs are involved in the degradation of several hydrocarbons including 1,3-dihalo-2-propanol, epichlorohydrin, 9,10-epoxy fatty acids, trans-2,3-epoxysuccinate and 2,3-chlorostyrene oxides (van der Werf *et al.*, 1998; Fretland and Omiecinski, 2000). The presence of EHs has been reported in bacteria recovered

from gasoline and oil-contaminated marine sediments (Kwon *et al.*, 2007; 2010; Woo *et al.*, 2007; 2013). Moreover, haloalkane dehalogenases (HDs) (EC 3.8.1.5) have attracted considerable attention due to their unique catalytic mechanism, broad substrate specificity, stability, enantioselectivity and catalytic efficiency (Koudelakova *et al.*, 2013). The HDs catalyse the cleavage of carbon–halogen bonds, which is a key step in the aerobic mineralization of many halogenated pollutants, such as oil compounds (Janssen *et al.*, 2005). Previous studies indicated HDs to be important for the preparation of optically pure building blocks for organic synthesis, recycling of by-products from chemical processes, decontamination of chemical warfare agents and for bio-sensing of environmental pollutants and protein tagging for cell imaging and protein analysis (Koudelakova *et al.*, 2013).

The current study aimed at bioprospection of shotgun sequence datasets generated from four mangrove soils for α/β -hydrolase fold proteins by using specific LED. In addition, a metagenomic fosmid library constructed from one oil-impacted mangrove site (BrMgv02) was used for sequence-based screening. The prevalence [relative abundance (RA)] of EHs and HDs was addressed, with special reference to the oil contamination, biodegradation and future potential industrial application.

Results and discussion

In this study, an analysis of the composition and diversity of metagenomic sequences encoding α/β -hydrolase fold proteins in four distinct mangrove soils was performed. Briefly, mangrove soil samples were collected in July 2008 in three distinct mangroves in the state of São Paulo (Brazil). Samples were divided in four groups: BrMgv01 and BrMgv02 (23°53'49"S; 46°12'28"W – Bertioga city) are two sites in the same mangrove separated by a small stream. This area has been affected by oil contamination (petroleum). Sample BrMgv03 (23°54'06"S; 45°15'03"W – Bertioga city) was taken from a site adjacent to BrMgv01-02 that had not been affected by oil, but by household waste. Finally, BrMgv04 (25°05'02"S; 47°57'42"W – Cananéia city) represents a sample from a pristine mangrove (Fig. 1) as detailed in Andreote and colleagues (2012). In total, 1.8 g (six samples per site) of soil from each area was subjected to total genomic DNA extraction, after which the DNAs were subjected to shotgun sequencing using the 454 GS-FLX titanium technology (Indianapolis, IN, USA). The sequences obtained (905 521 unassembled sequences with an average length of 236 bp) were sorted and trimmed based on length and quality, using an in-house python script (Jiménez *et al.*, 2012b). The total numbers of trimmed sequences obtained for each mangrove area were 249 993 for BrMgv01 (average read length ~ 235 bp), 231 233 for BrMgv02

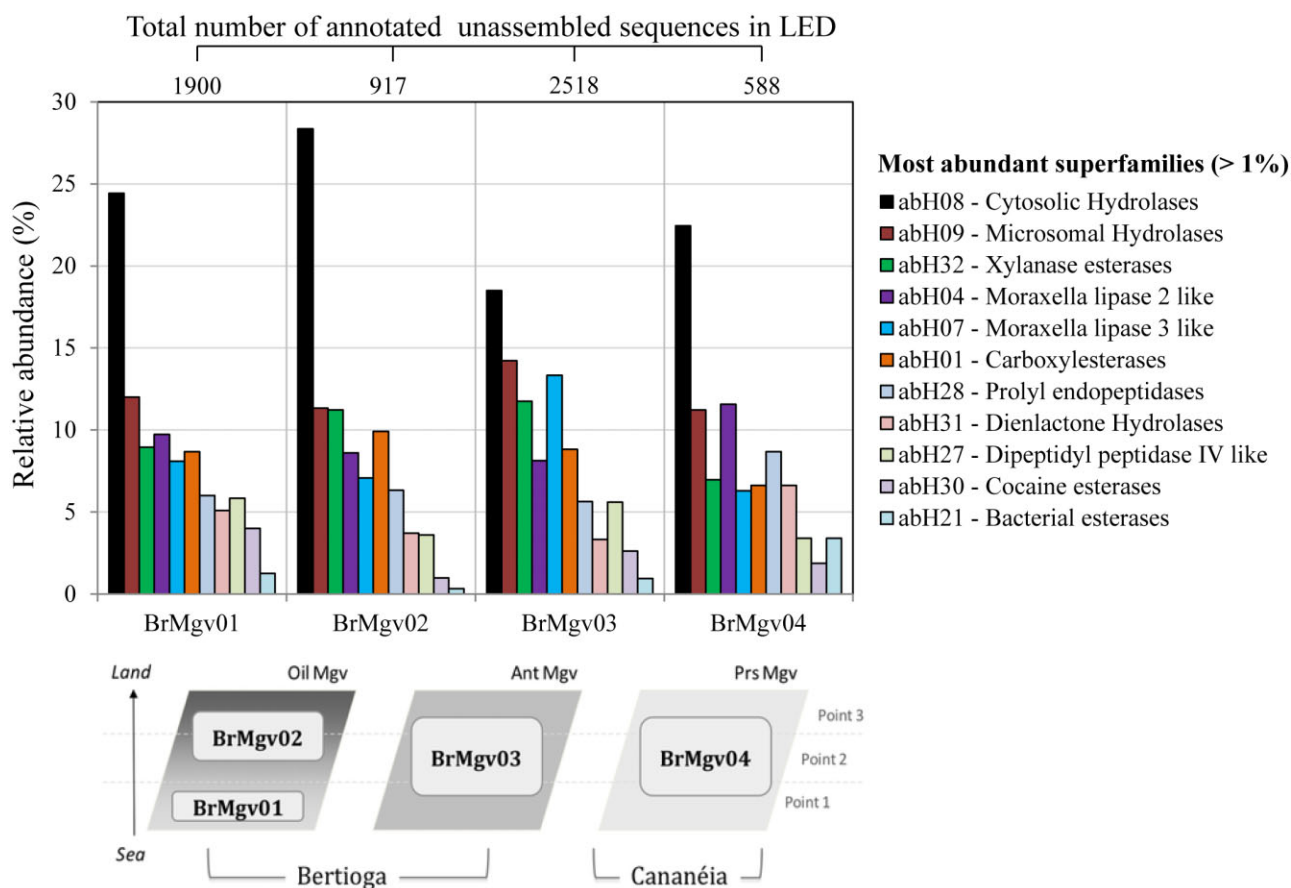


Fig. 1. Relative abundance (%) of the most abundant α/β -hydrolase fold protein superfamilies in the four mangrove metagenomes (BrMgv01 to BrMgv04). Oil Mgv: oil-polluted mangrove site; Ant Mgv: anthropogenic polluted site; Prs: pristine mangrove site. The database was downloaded from the LED ftp website – <http://www.led.uni-stuttgart.de/> – release 3.0, last update on 10 December 2009. Normalization was performed using the total number of annotated sequences (in LED) in each dataset.

(~ 238 bp), 214 921 for BrMgv03 (~ 248 bp) and 217 605 for BrMgv04 (~ 223 bp). These sequences were uploaded to the metagenomic RAST (MG-RAST) server and made publicly accessible under the project codes 4451033.3, 4451034.3, 4451035.3 and 4451036.3 for mangroves BrMgv01, BrMgv02, BrMgv03, and BrMgv04 respectively. As a complement to a previous study (Andreote *et al.*, 2012), we performed BLASTX against LED (Fischer and Pleiss, 2003) using a cut-off e-value of $1e^{-5}$, as in other studies in which moderately rigid criteria were used to find genetic novelty and to evaluate functional and taxonomic profiles (Jung *et al.*, 2011; Jiménez *et al.*, 2012b; Mendes *et al.*, 2014). It is important to note that BLASTX has been successfully used against LED in other studies (Kim *et al.*, 2009; Damon *et al.*, 2012). However, we are aware of the fact that these parameter settings may result in spurious hits and thus data need to be carefully re-examined. To address this critical issue, we performed manual annotation in the best hits. With this strategy, the information retrieved was compared across the four datasets. Thus, totals of 1900 (0.8% RA), 917 (0.4% RA), 2518 (1.1% RA)

and 588 (0.2% RA) unassembled sequences were found to match 30 different α/β -hydrolase superfamilies, for mangrove samples BrMgv01, BrMgv02, BrMgv03 and BrMgv04 respectively (Fig. 1). Collectively, these sequences matched 22 superfamilies in the class GGGX. In the LED, proteins were assigned to the classes GX, GGGX and Y, in accordance with their amino acid sequences and the structures of the oxy anion holes (Pleiss *et al.*, 2000). The oxy anion hole helps to stabilize the negatively charged transition state that occurs during enzymatic hydrolysis (Nardini and Dijkstra, 1999).

The most abundant superfamilies across all metagenomes were annotated as cytosolic hydrolases (abH08) (~ 23% RA) and microsomal hydrolases (abH09) (~ 12% RA) (except in BrMgv04), followed by *Moraxella* lipase 2-like sequences (abH04) (9.7% RA in BrMgv01 and 11.5% RA in BrMgv04), xylanase esterases (abH32) (11.7% RA in BrMgv02) and *Moraxella* lipase 3-like (abH07) (13.3% RA in BrMgv03) (Fig. 1). Superfamily abH08 consists of 15 homologous families that include 3188 protein entries in LED. Into the abH08 superfamily, we can found a large

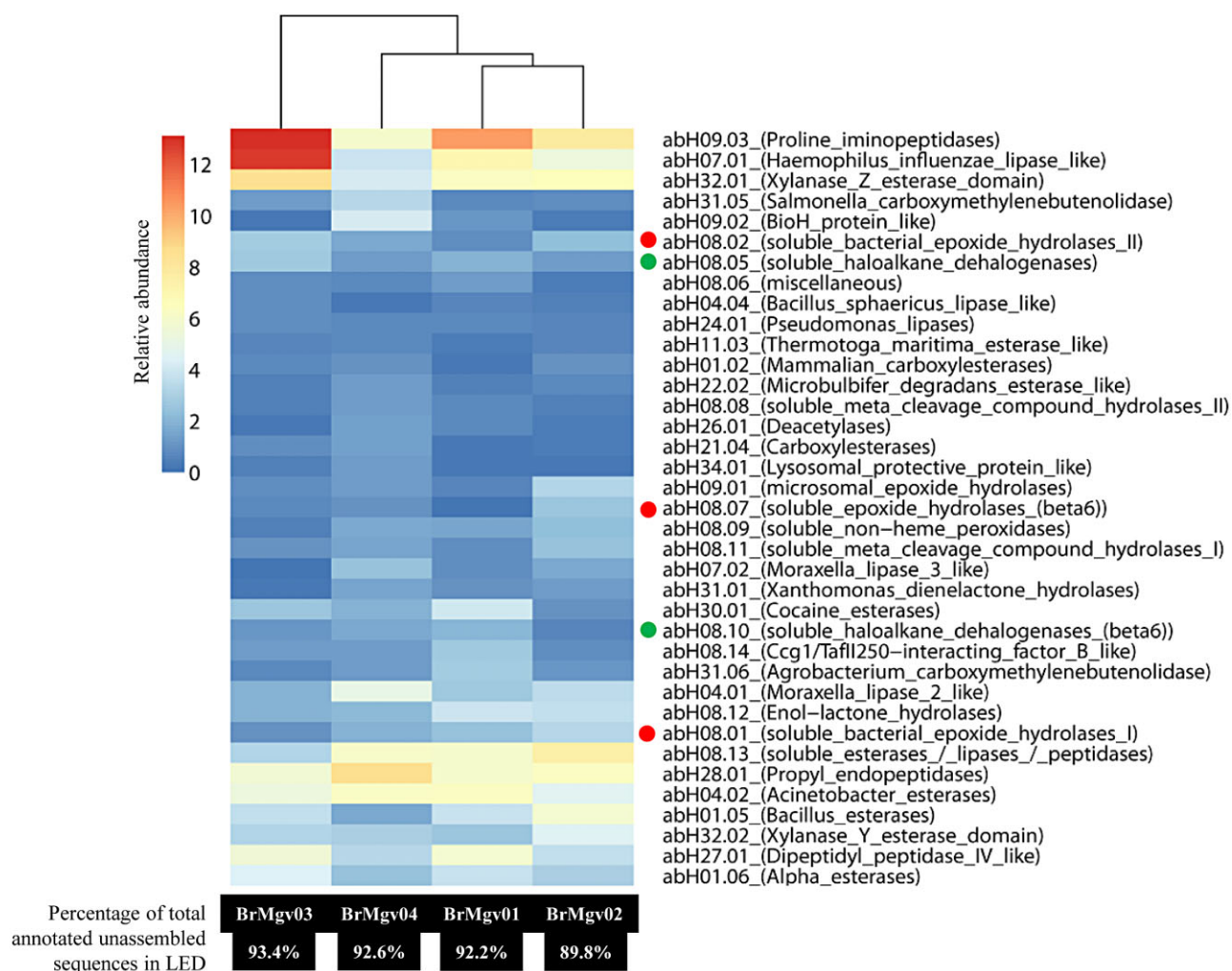


Fig. 2. Relative abundance (%) of the most abundant α/β -hydrolase fold protein families in four mangrove metagenomes (BrMgv01–BrMgv04). Hierarchical dendrogram and heat-mapping were performed using cluster V3.0 (Eisen *et al.*, 1998) software. Red and green dots represent the epoxide hydrolase (EHs) and haloalkane dehalogenase (HDs) families respectively.

group of bacterial EHs, non-heme peroxidases and HDs. In addition, superfamily abH09 contains three families, namely microsomal EHs, BioH protein like (biotin biosynthesis) and proline iminopeptidases (Fischer and Pleiss, 2003). Furthermore, proline iminopeptidases, propyl endopeptidases and dipeptidyl peptidases make part of the bacterial proteolytic system, which has been reported as being key for bacterial nitrogen utilization (nitrogen from amino acids), especially under nitrogen-limiting conditions (Kunji *et al.*, 1996; Li *et al.*, 2010). Nitrogen limitation may reign in anoxic soil mangroves, in which denitrification leading to gaseous nitrogen effluxes is rampant (Fernandes *et al.*, 2012).

We then assessed the distribution of the unassembled sequences over the predicted protein families per mangrove sample. The prevalences of proline iminopeptidases (abH09.03), *Haemophilus influenzae* lipase-like (abH07.01), xylanase Z esterase domain (abH32.01) and

dipeptidyl peptidases (abH27.01) were highest in BrMgv01 (10.3%, 7.2%, 6.4% and 5.8% RA respectively) and BrMgv03 (13.1%, 12.7%, 8.5% and 5.5% RA respectively). The families encompassing *Acinetobacter* esterases (abH04.02) and propyl endopeptidases (abH28.01) showed highest prevalences (6.3 and 8.6% RA respectively) in the unpolluted mangrove site (BrMgv04). Conversely, EHs, belonging to families abH08.01, abH09.01, abH08.07 and abH08.02, were abundant in BrMgv02 (3.3%, 3.1%, 2.6% and 2.3% RA respectively) compared with the unpolluted mangrove BrMgv04 (1.8%, 1.1%, 1.0% and 1.7% RA respectively) (Fig. 2). On another matter, genes encoding xylanases/esterases were very prevalent in all mangrove samples (between 4% and 8% RA), irrespective of pollution. This suggests that the conversion of hemicellulose is likely to occur in these environments, which are anaerobic most of the time (Benner *et al.*, 1984; Ye *et al.*, 2010).

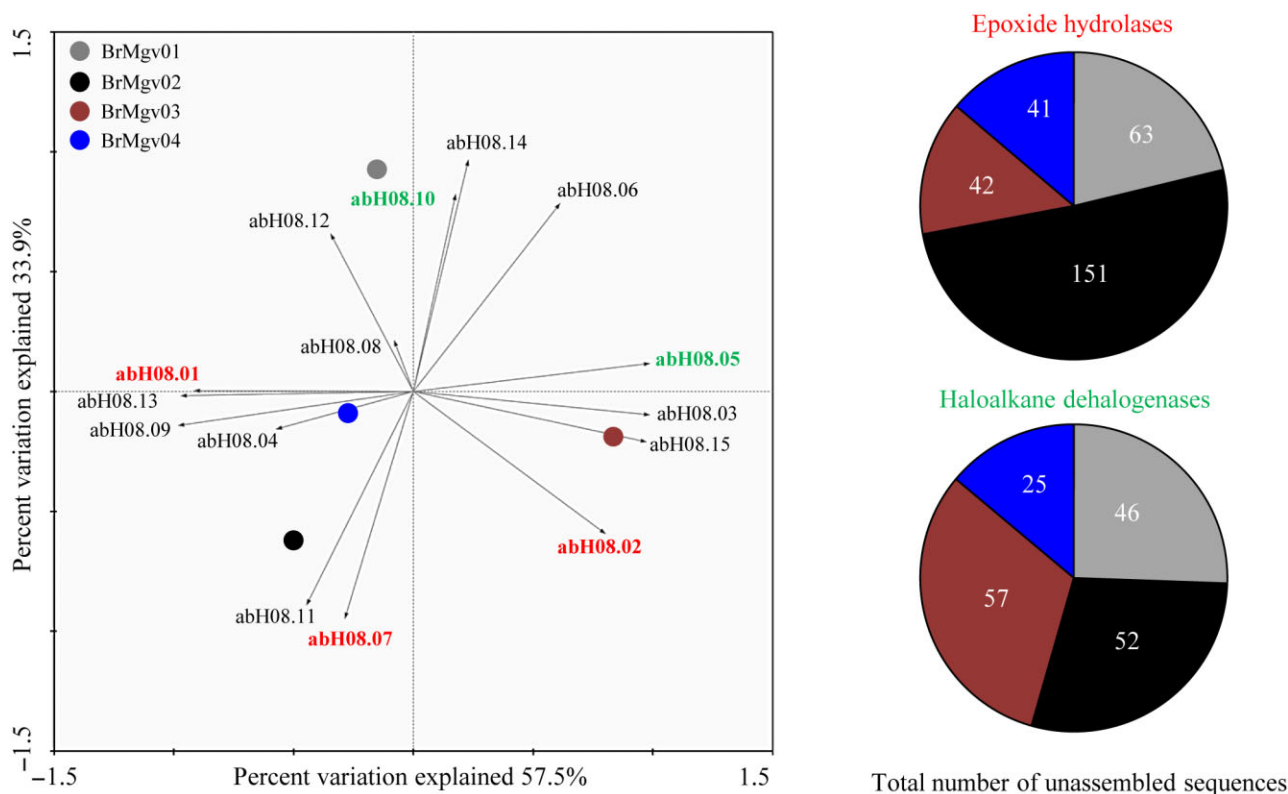


Fig. 3. Left: principal component analysis (PCA) within the family abH08 using Canoco software v4.52 (Wageningen, the Netherlands). Right: Number of unassembled sequences annotated in the abH08.02 [epoxide hydrolases (EHs)] and abH08.05 [haloalkane dehalogenases (HDs)] families in BrMgv samples.

To shed light on the α/β -hydrolase fold protein superfamilies detected in our dataset, we focused on the protein superfamily that was most abundant across the board, i.e. abH08. Principal component analyses showed that EHs, families abH08.07 and abH08.02, were preferentially present in BrMgv02 and BrMgv03 respectively (Fig. 3). In terms of numbers of annotated sequences, the prevalences of these predicted enzymes (abH08.01, abH08.07 and abH08.02) were highest in BrMgv02 site (151 unassembled sequences – 30% RA based on the number of annotated sequences within abH08 family) comparatively to BrMgv04 (41 unassembled sequences – 17% RA). Regarding HDs (abH08.10 and abH08.05), highest numbers of annotated genes were observed in the polluted mangroves BrMgv02 (52 unassembled sequences – 11% RA) and BrMgv03 (57 unassembled sequences – 22% RA) (Fig. 3). Interestingly, van Loo and colleagues (2006) reported the screening of various genomic databases for the presence of EHs to find ways to express these proteins in different bacterial hosts. In addition, putative open reading frames for EHs and subsequent expression have been investigated in *Cupriavidus metallidurans* CH34 (Kumar *et al.*, 2011). Clearly, novel EHs, provided they offer features such as enhanced activity, substrate specificity and/or stability

in the face of chemicals, are useful in the production of a range of compounds (e.g. β 3-adrenergic receptor agonists, anti-obesity and anti-inflammatory drugs, nematocides, anticancer agents, anti-fungal chemicals). In addition, they may also serve the detoxification of xenobiotics such as polyaromatic hydrocarbons and the production of enantiopure epoxides and vicinal diols from cheap racemic epoxides (Lee and Shuler, 2007).

In the light of the high prevalence of sequences related to EHs (family abH08.02) and HDs (family abH08.05) in highly oil-impacted mangrove BrMgv02 (73 and 36 unassembled sequences respectively), this specific dataset and the best BLAST hits (based on e-values) were further analysed. For the former unassembled sequences, we found some putative novel enzymes (based on relatively low amino acid identity: < 60%) that were mostly affiliated with similar proteins (amino acid identity between 50% and 73%) from *Streptomyces*, *Hyphomonas*, *Bradyrhizobium* and *Phenylobacterium*. For the abH08.05 family, HDs were affiliated to proteins from *Salinispora*, *Alcanivorax*, *Photobacterium*, *Moritella* and *Chloroflexus* (Table 1 – approach 1). It is important to stress that the taxonomic affiliation is only an indication of identity as it is based on function of the most similar or homologous proteins, and thus may not directly reflect the

Table 1. Putative epoxide hydrolases (EHs) retrieved from the BrMgv02 dataset.

Approach (Mangrove)	Best-hit protein (NCBI Accession No.)	Putative microbial source based on homology (Best hit protein)	Hit-length (aa)	Query-length (bp)	e-value	Amino acid identity (%)	ESTHER ^c
1 (BrMgv02)	Putative epoxide hydrolase (YP_761108)	<i>Hyphomonas neptunium</i>	320	345	7e-48	73.04	XE
	Epoxide hydrolase (NP_823281)	<i>Streptomyces avermitilis</i>	328	394	2e-28	48.46 ^a	XE
	Epoxide hydrolase (YP_002203910)	<i>Streptomyces sviveus</i>	330	393	8e-28	47.69 ^a	XE
	Epoxide hydrolase (NP_767754)	<i>Bradyrhizobium diazoefficiens</i>	330	345	7e-24	50 ^a	XE
	Epoxide hydrolase (YP_002129963.1)	<i>Phenylobacterium zucineum</i>	321	347	4e-23	51 ^a	XE
2 (BrMgv02)	Alpha/beta hydrolase fold protein (WP_008687080.1)	<i>Rhodopirellula sallentina</i>	319	850	1e-65	40 ^{ab}	XE
	Epoxide hydrolase – like protein (YP_003705817.1)	<i>Frankia</i> sp.	398	351	4e-32	51 ^{ab}	XE
	Alpha/beta hydrolase fold protein (YP_002544714.1)	<i>Truepera radiovictrix</i>	292	504	3e-38	69 ^b	XE
	Alpha/beta hydrolase fold protein (YP_003336243.1)	<i>Streptosporangium roseum</i>	307	1459	3e-75	55 ^{ab}	XE
	Epoxide hydrolase domain protein (WP_007534516.1)	<i>Rhizobium mesoamericanum</i>	199	345	3e-21	49 ^{ab}	XE
	Epoxide hydrolase (WP_009025691)	<i>Bradyrhizobium</i> sp.	334	802	8e-74	46 ^{ab}	XE
	Alpha/beta hydrolase (YP_727726.1)	<i>Rhodospirillum rubrum</i>	305	1039	4e-116	64 ^b	XE

a. Putative novel protein based on low amino acid identity (< 60%).

b. Complete genes.

c. Classification based on BLASTX against ESTHER database (Lenfant *et al.*, 2013).

The table was constructed using two different approaches: (1) Whole metagenome sequencing – and BLASTX against the LED; (2) Fosmid library sequencing – and BLASTX against the NCBI using the MG-RAST pipeline.

XE, Block_X, Family Epoxide_hydrolase.

true microbial source. Thus a whole new suite of partial genes encoding novel enzymes within these two classes was unlocked. These are hypothesized to function best under the environmental conditions of the habitat, i.e. the local salt, oxic/anoxic conditions as well as the presence of hydrocarbons (Arulazhagan and Vasudevan, 2011; Arfi *et al.*, 2013). Such enzymes might well suit industrial and environmental (pollutant removal) needs, as will be explored in future work. According to Andreote and colleagues (2012), the bacterial community at the BrMgv02 mangrove site were dominated by *Proteobacteria* belonging to the classes *Delta* < *Gamma* < *Alpha*. Moreover, *Alphaproteobacteria*, next to *Actinobacteria* were significantly raised compared with BrMgv04. Hence, we surmised that such bacterial groups were carriers of the biodegradative functions in the oil-polluted BrMgv02, which is consistent with findings in other estuarine ecosystems (Greer, 2010).

In a second stage, we studied selected sequences retrieved from a fosmid library constructed using the CopyControl Fosmid Library Production Kit – Epicentre which contains the pCC2FOS as a vector and *Escherichia coli* EPI 300 T1R Phage T1 resistant [F[−] mcrA Δ(*mrr-hsdRMS-mcrBC*) (Str^R) φ80dlacZΔM15 ΔlacX74 *recA1 endA1 araD139Δ(ara, leu)7697 galU galK λ[−] rpsL nupG*

trfA tonA dhfrJ] as host cells. Environmental DNA extracted from the oil-impacted mangrove soil BrMgv02 was used to construct the metagenomic library, with insert sizes ranging from 25 to 40 kb. A total of 12 900 clones was obtained and further subjected to 454 pyrosequencing (GS FLX Titanium technology), yielding a total of 1 380 509 unassembled sequences with average length of 484 bp (approximately 624 Mb). The sequences were assembled into contigs using CLC Genomics Workbench (version 6.5.1; CLC Bio, Cambridge, MA, USA) (under default parameters) resulting in 118 882 sequences (average length 913 bp), totalling 108 Mb. The data were annotated using the MG-RAST pipeline (cut-off e-value of 1e-5) (Meyer *et al.*, 2008). This allowed the detection of seven putative EHs in fosmids (fosmid library metagenomic data were deposited at MG-RAST under the ID No. 4555913.3). These were further compared and were found to be affiliated with EHs in genomes of *Rhodopirellula*, *Bradyrhizobium* (family abH08.02), *Truepera*, *Streptosporangium* (family abH08.01), *Rhizobium*, *Frankia* (family abH09.01) and *Rhodospirillum* (family abH08.07) (Table 1 – approach 2). Moreover, five HDs sequences were detected, and these matched those present in the genomes of *Anaeromyxobacter*, *Oceanicaulis* (family abH08.05), *Mycobacterium* (families abH08.05 and

Table 2. Putative haloalkane dehalogenases (HDs) retrieved from the BrMgv02 dataset.

Approach (Mangrove)	Best-hit protein (NCBI Accession No.)	Putative microbial source based on homology (Best hit protein)	Hit-length (aa)	Query-length (bp)	e-value	Amino acid identity (%)	ESTHER ^c
1 (BrMgv02)	Haloalkane dehalogenase (YP_001537995)	<i>Salinispora arenicola</i>	310	412	1e-29	65.8	X1
	Haloalkane dehalogenase (YP_694135)	<i>Alcanivorax borkumensis</i>	296	410	5e-29	63.5	X1
	Putative haloalkane dehalogenase (ZP_01221858)	<i>Photobacterium profundum</i>	303	242	1e-27	62.5	X1
	Putative haloalkane dehalogenase (ZP_01897865)	<i>Moritella</i> sp.	292	190	2e-25	75.8	X1
	Haloalkane dehalogenase (ZP_02986139)	<i>Chloroflexus</i> sp.	293	306	5e-19	45 ^a	X1
2 (BrMgv02)	Haloalkane dehalogenase (YP_001377212.1)	<i>Anaeromyxobacter</i> sp.	304	840	8e-37	49 ^{ab}	X1
	Haloalkane dehalogenase (WP_020727623.1)	<i>Mycobacterium marinum</i>	325	1752	1e-146	80 ^b	X2
	Haloalkane dehalogenase (WP_022699531.1)	<i>Oceanicaulis alexandrii</i>	301	249	1e-46	64 ^b	X1
	Haloalkane dehalogenase (YP_003336243.1)	<i>Mycobacterium vanbaalenii</i>	298	607	5e-81	61 ^b	X1
	Haloalkane dehalogenase (WP_007078410.1)	<i>Aeromicrobium marinum</i>	290	1241	3e-65	52 ^{ab}	X1

a. Putative novel protein based on low amino acid identity (< 60%).

b. Complete genes.

c. Classification based on BLASTX against ESTHER database (Lenfant *et al.*, 2013).

The table was constructed using two different approaches: (1) Whole metagenome sequencing – and BLASTX against the LED; (2) Fosmid library sequencing – and BLASTX against the NCBI using the MG-RAST pipeline.

X1, Block_X, Family Haloalkane_dehalogenase-HLD1; X2, Block_X, Family Haloalkane_dehalogenase-HLD2.

abH08.10) and *Aeromicrobium* (abH08.10). In some cases (five EHs and two HDs), the sequence identity with already described genes (database) was low (< 60%), suggesting that the sequences could represent putative new proteins. The hierarchical classification of these 22 (complete and/or partial) genes, based on the ESTHER database, was done (Tables 1 and 2). Importantly, the activity of such enzymes may not be detected in *E. coli* host cells, specially due to the different expression systems present in this species when compared with those of *Actinobacteria*, *Alphaproteobacteria*, *Deinococcus-Thermus* and *Planctomycetes* members (Gabor *et al.*, 2004). On the other hand, the use of degenerate primers has been applied as a sequence-driven approach to identify enzymes directly from metagenomes. For example, Kotik and colleagues (2009) amplified fragments of EH genes using degenerate primers targeted to conserved motifs, followed by assembly by genome walking. These results highlight the importance of using both sequence and function-based approaches in metagenomic library screenings, thus circumventing problems inherent to either lack of heterologous expression efficiency or limited sequence information based on known protein sequences available at public databases. Our fosmid-based approach opens up the possibility of finding whole operons, start/stop codons and expression signals on the basis of the genetic information gathered in this study. The EHs and HDs encoding genes from the oil-impacted mangrove soils can thus be

further expressed in appropriate expression vectors, for future practical use in industrial or biotechnological processes. Comparison of the EHs genes between the BrMgv02 metagenome and the fosmid library datasets revealed protein families (abH08.01, abH09.01, abH08.07 and abH08.02) to be coincident. Similar results were found with HDs, as the main protein families found within the BrMgv02 metagenome were also observed in the fosmid library dataset (abH08.05 and abH08.10). Thus, the metagenomic library represented to a considerable extent the diversity of EHs and HDs present in the analysed mangrove sample.

Clearly, *Actinobacteria* and *Alphaproteobacteria* might serve as genetic sources for EHs bio-exploration. For instance, analysis of the genome of the actinobacterium *Mycobacterium tuberculosis* revealed an unusually large number of potential EHs, i.e. nine EHs genes occurred scattered on the genome (Johansson *et al.*, 2005). Also, the presence of EHs in *Agrobacterium radiobacter* AD1 has been reported (Rink *et al.*, 1997). These were further engineered towards an increasing activity for industrial purposes (Rui *et al.*, 2005). Protein engineering proved to be an efficient method to tailor α/β -hydrolase fold enzymes towards a desired property. Moreover, enzymes with completely new catalytic activities have been generated, for instance the conversion of an esterase from *Pseudomonas fluorescens* into an EH (Jochens *et al.*, 2009). The EHs, HDs and halo-peroxidases have a typical

lipase catalytic triad (G-X-S-X-G) and share approximately 25% of amino acid identity with lipases belonging to the family V (Arpigny and Jaeger, 1999; Tirawongsaroj *et al.*, 2008). In the catalytic triad, the nucleophilic aspartate carries out an attack on the carbon atom of the epoxide ring, thus displacing the oxygen and producing a covalent intermediate compound (de Vries and Janssen, 2003). Soluble EHs have recently been found in an Andean forest soil metagenome, in this case affiliated to the bacterium *Streptomyces scabies* (Montaña *et al.*, 2012). In addition, Procópio and colleagues (2013) reported the presence of five putative genes encoding EHs in the genome of *Dietzia cinnamomea* (a common soil *Actinobacterium*). These studies are consistent with the notion that members of the *Actinobacteria* can produce EHs in the environment, which supports their potential use as bioremediation agents, for instance in oil-contaminated systems. Moreover, we also found HDs belonging to *Alcanivorax* and *Phenylobacterium*. These microorganisms are known as 'hydrocarbonoclastic' based on their capacity to degrade an exceptionally broad range of haloalkane hydrocarbons (Sabirova *et al.*, 2006; dos Santos *et al.*, 2011). The genus *Phenylobacterium* (a facultatively anaerobic bacterium) has a unique preference for phenyl moieties from heterocyclic compounds such as chloridazon, antipyrine and pyramidon (Oh and Roh, 2012). Conversely, *Alcanivorax* species has also been reported as a key bacterial group present in crude oil enrichments based on mangrove soils as the microbial source (Brito *et al.*, 2006).

Conclusions

Current bottlenecks in high-throughput metagenome analysis are mostly due to problems related to sequence annotation. This factor can drastically affect the interpretation of a given dataset, especially in the case of enzyme annotation (Hoff, 2009; Schnoes *et al.*, 2009). In this sense, it becomes important to use specific and curated databases, which – in combination with manual annotation – can improve our capability of data mining. In this study, we make use of metagenomics datasets from mangrove soils to investigate the prevalence and diversity of genes for α/β -hydrolase fold related proteins, using a specific database. Sequences, predicted to belong to *Actinobacteria*, *Chloroflexi*, *Deinococcus-Thermus*, *Planctomycetes* and *Proteobacteria* EHs or HDs codifying genes are described and analysed. Moreover, the description of the EHs and HDs will be further explored in the context of the bioconversion of hydrocarbons in oil-contaminated environments. Our results might represent a first step towards the development of a totally synthetic metagenomics approach (synthesis-cloning-expression), to be broadly applied in mangrove tropical ecosystems.

Finally, we conclude that the presence of hydrocarbons in mangrove soils has an effect on the abundance and diversity of α/β -hydrolase fold proteins, which were mostly heightened in EHs and HDs.

Acknowledgements

We thank C. Guzman, D. Riaño and S. Restrepo from Andes University (Bogotá D.C, Colombia) for their support through bioinformatics analysis.

Conflict of interest

The authors declare no conflict of interest.

References

- Andreote, F.D., Jiménez, D.J., Chaves, D., Dias, A.C., Luvizotto, D.M., Dini-Andreote, F., *et al.* (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS ONE* **7**: e38600.
- Arfi, Y., Chevret, D., Henrissat, B., Berrin, J.G., Levasseur, A., and Record, E. (2013) Characterization of salt-adapted secreted lignocellulolytic enzymes from the mangrove fungus *Pestalotiopsis* sp. *Nat Commun* **4**: 1810.
- Arpigny, J.L., and Jaeger, K.E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J* **343**: 177–183.
- Arulazhagan, P., and Vasudevan, N. (2011) Biodegradation of polycyclic aromatic hydrocarbons by a halotolerant bacterial strain *Ochrobactrum* sp. VA1. *Mar Pollut Bull* **62**: 388–394.
- Barth, S., Fischer, M., Schmid, R.D., and Pleiss, J. (2004) The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. *Bioinformatics* **20**: 2845–2847.
- Benner, R., Maccubbin, A.E., and Hodson, R.E. (1984) Anaerobic biodegradation of the lignin and polysaccharide components of lignocellulose and synthetic lignin by sediment microflora. *Appl Environ Microbiol* **47**: 998–1004.
- Brito, E.M., Guyoneaud, R., Goñi-Urriza, M., Ranchou-Peyruse, A., Verbaere, A., Crapez, M.A., *et al.* (2006) Characterization of hydrocarbonoclastic bacterial communities from mangrove sediments in Guanabara Bay, Brazil. *Res Microbiol* **157**: 752–762.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37** (Database issue): D233–D238.
- Chistoserdova, L. (2010) Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* **32**: 1351–1359.
- Choi, W.J. (2009) Biotechnological production of enantiopure epoxides by enzymatic kinetic resolution. *Appl Microbiol Biotechnol* **84**: 239–247.
- Couto, G.H., Glogauer, A., Faoro, A., Chubatsu, L.S., Souza, E.M., and Pedrosa, F.O. (2010) Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genet Mol Res* **9**: 514–523.

- Damon, C., Lehenbre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., and Marmesse, R. (2012) Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PLoS ONE* **7**: e28967.
- Dias, A.C.F., Andreote, F.D., Dini-Andreote, F., Lacava, P.T., Sá, A.L.B., Melo, I.S., *et al.* (2009) Diversity and biotechnological potential of culturable bacteria from Brazilian mangrove sediment. *World J Microbiol Biotechnol* **25**: 1305–1311.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* **95**: 14863–14868.
- Ekkers, D.M., Cretoiu, M.S., Kielak, A.M., and van Elsas, J.D. (2012) The great screen anomaly—a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* **93**: 1005–1020.
- Fawal, N., Li, Q., Savelli, B., Brette, M., Passaia, G., Fabre, M., *et al.* (2013) PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res* **41** (Database issue): D441–D444.
- Fernández-Arrojo, L., Guazzaroni, M.E., López-Cortés, N., Belouqui, A., and Ferrer, M. (2010) Metagenomic era for biocatalyst identification. *Curr Opin Biotechnol* **21**: 725–733.
- Fernandes, S.O., Michotey, V.D., Guasco, S., Bonin, P.C., and Bharathi, P.A. (2012) Denitrification prevails over anammox in tropical mangrove sediments (Goa, India). *Mar Environ Res* **74**: 9–19.
- Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res* **31**: 319–321.
- Fretland, A.J., and Omiecinski, C.J. (2000) Epoxide hydrolases: biochemistry and molecular biology. *Chem Biol Interact* **129**: 41–59.
- Gabor, E.M., Alkema, W.B., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**: 879–886.
- Greer, C.W. (2010) Bacterial diversity in hydrocarbon-polluted rivers, estuaries and sediments. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, K.N. (ed.). Berlin Heidelberg, Germany: Springer-Verlag, pp. 2329–2338.
- Hoff, K.J. (2009) The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics* **10**: 520.
- Holmquist, M. (2000) Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr Protein Pept Sci* **1**: 209–235.
- Janssen, D.B., Dinkla, I.J., Poelarends, G.J., and Terpstra, P. (2005) Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ Microbiol* **7**: 1868–1882.
- Jiménez, D.J., Montaña, J.S., Alvarez, D., and Baena, S. (2012a) A novel cold active esterase derived from Colombian high Andean forest soil metagenome. *World J Microbiol Biotechnol* **1**: 361–370.
- Jiménez, D.J., Andreote, F.D., Chaves, D., Montaña, J.S., Osorio-Forero, C., Junca, H., *et al.* (2012b) Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. *PLoS ONE* **7**: e25069.
- Jochens, H., Stiba, K., Savile, C., Fujii, R., Yu, J.G., Gerassenskov, T., *et al.* (2009) Converting an esterase into an epoxide hydrolase. *Angew Chem Int Ed Engl* **48**: 3532–3535.
- Johansson, P., Unge, T., Cronin, A., Arand, M., Bergfors, T., Jones, T.A., and Mowbray, S.L. (2005) Structure of an atypical epoxide hydrolase from *Mycobacterium tuberculosis* gives insights into its function. *J Mol Biol* **351**: 1048–1056.
- Jung, J.Y., Lee, S.H., Kim, J.M., Park, M.S., Bae, J.W., Hahn, Y., *et al.* (2011) Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl Environ Microbiol* **77**: 2264–2274.
- Kathiresan, K., and Bingham, B.L. (2001) Biology of mangroves and mangrove ecosystems. *Adv Mar Biol* **40**: 81–251.
- Kim, E.Y., Oh, K.H., Lee, M.H., Kang, C.H., Oh, T.K., and Yoon, J.H. (2009) Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microbiol* **75**: 257–260.
- Kotik, M., Stepanek, V., Maresova, H., Kyslik, P., and Archelas, A. (2009) Environmental DNA as a source of novel epoxide hydrolase reacting with aliphatic terminal epoxides. *J Mol Catal B-Enzym* **56**: 288–293.
- Koudelakova, T., Bidmanova, S., Dvorak, P., Pavelka, A., Chaloupkova, R., Prokop, Z., and Damborsky, J. (2013) Haloalkane dehalogenases: biotechnological applications. *Biotechnol J* **8**: 32–45.
- Kourist, R., Jochens, H., Bartsch, S., Kuipers, R., Padhi, S.K., Gall, M., *et al.* (2010) The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *ChemBiochem* **11**: 1635–1643.
- Kumar, R., Wani, S.I., Chauhan, N.S., Sharma, R., and Sareen, D. (2011) Cloning and characterization of an epoxide hydrolase from *Cupriavidus metallidurans*-CH34. *Protein Expr Purif* **79**: 49–59.
- Kunji, E.R.S., Mierau, I., Hagfing, A., Poolman, B., and Konings, W.N. (1996) The proteolytic systems of lactic acid bacteria. *A van Leeuw J Microb* **70**: 187–221.
- Kwon, K.K., Woo, J.H., Yang, S.H., Kang, J.H., Kang, S.G., Kim, S.J., *et al.* (2007) *Altererythrobacter epoxidivorans* gen. nov., sp. nov., an epoxide hydrolase-active, mesophilic marine bacterium isolated from cold-seep sediment, and reclassification of *Erythrobacter luteolus* Yoon *et al.* 2005 as *Altererythrobacter luteolus* comb. nov. *Int J Syst Evol Microbiol* **57** (Part 10): 2207–2211.
- Kwon, T.H., Kim, J.T., and Kim, J.S. (2010) Application of a modified sublimation method to screen for PAH-degrading microorganisms. *Korean J Microbiol* **46**: 109–111.
- Lee, E.Y., and Shuler, M.L. (2007) Molecular engineering of epoxide hydrolase and its application to asymmetric and enantioconvergent hydrolysis. *Biotechnol Bioeng* **98**: 318–327.
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., and Chatonnet, A. (2013) ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res* **41** (Database issue): D423–D429.

- Li, N., Wu, J.M., Zhang, L.F., Zhang, Y.Z., and Feng, H. (2010) Characterization of a unique proline iminopeptidase from white-rot basidiomycetes *Phanerochaete chrysosporium*. *Biochimie* **92**: 779–788.
- van Loo, B., Kingma, J., Arand, M., Wubbolts, M.G., and Janssen, D.B. (2006) Diversity and biocatalytic potential of epoxide hydrolases identified by genome analysis. *Appl Environ Microbiol* **72**: 2905–2917.
- Mendes, L.W., Kuramae, E.E., Navarrete, A.A., van Veen, J.A., and Tsai, S.M. (2014) Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J* **8**: 1577–1587.
- Meyer, F., Paarmann, D., Souza, M.D., Olson, R., Glass, E.M., Kubal, M., *et al.* (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.
- Montaña, J.S., Jiménez, D.J., Hernández, M., Ángel, T., and Baena, S. (2012) Taxonomic and functional assignment of cloned sequences from high Andean forest soil metagenome. *A van Leeuw J Microb* **101**: 205–215.
- Nardini, M., and Dijkstra, B.W. (1999) Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* **9**: 732–737.
- Oh, Y.S., and Roh, D.H. (2012) *Phenylobacterium muchangponense* sp. nov., isolated from beach soil, and emended description of the genus *Phenylobacterium*. *Int J Syst Evol Microbiol* **62** (Part 4): 977–983.
- Pleiss, J., Fischer, M., Peiker, M., Thiele, C., and Schmid, R.D. (2000) Lipase Engineering Database—understanding and exploiting sequence-structure-function relationships. *J Mol Catal B-Enzym* **10**: 491–508.
- Procópio, L., Macrae, A., van Elsas, J.D., and Seldin, L. (2013) The putative α/β -hydrolases of *Dietzia cinnamon* P4 strain as potential enzymes for biocatalytic applications. *A van Leeuw J Microb* **103**: 635–646.
- Rink, R., Fennema, M., Smids, M., Dehmelt, U., and Janssen, D.B. (1997) Primary structure and catalytic mechanism of the epoxide hydrolase from *Agrobacterium radiobacter* AD1. *J Biol Chem* **272**: 14650–14657.
- Rui, L., Cao, L., Chen, W., Reardon, K.F., and Wood, T.K. (2005) Protein engineering of epoxide hydrolase from *Agrobacterium radiobacter* AD1 for enhanced activity and enantioselective production of (R)-1-phenylethane-1,2-diol. *Appl Environ Microbiol* **71**: 3995–4003.
- Sabirova, J.S., Ferrer, M., Regenhardt, D., Timmis, K.N., and Golyshin, P.N. (2006) Proteomic insights into metabolic adaptations in *Alcanivorax borkumensis* induced by alkane utilization. *J Bacteriol* **188**: 3763–3773.
- Sahoo, K., and Dhal, N. (2009) Potential microbial diversity in mangrove ecosystems: a review. *IJMS* **38**: 249–256.
- dos Santos, H.F., Cury, J.C., do Carmo, F.L., dos Santos, A.L., Tiedje, J., van Elsas, J.D., *et al.* (2011) Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. *PLoS ONE* **6**: e16943.
- Sareen, D., and Kumar, R. (2011) Prospecting for efficient enantioselective epoxide hydrolases. *Indian J Biotech* **10**: 161–177.
- Schnoes, A.M., Brown, S.D., and Babbitt, P.C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* **5**: e1000605.
- Sirim, D., Wagner, F., Wang, L., Schmid, R.D., and Pleiss, J. (2011) The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database (Oxford)* **2011**: bar006.
- Thompson, C.E., Beys-da-Silva, W.O., Santi, L., Berger, M., Vainstein, M.H., Guima Rães, J.A., and Vasconcelos, A.T. (2013) A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. *AMB Express* **3**: 65.
- Tirawongsoj, P., Sriprang, R., Harnpicharnchai, P., Thongarama, T., Champreda, V., Tanapongpipat, S., *et al.* (2008) Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J Biotechnol* **133**: 42–49.
- de Vries, E.J., and Janssen, D.B. (2003) Biocatalytic conversion of epoxides. *Curr Opin Biotechnol* **14**: 414–420.
- van der Werf, M.J., Overkamp, K.M., and de Bont, J.A. (1998) Limonene-1,2-epoxide hydrolase from *Rhodococcus erythropolis* DCL14 belongs to a novel class of epoxide hydrolases. *J Bacteriol* **180**: 5052–5057.
- Woo, J.H., Hwang, Y.O., Kang, S.G., Lee, H.S., Cho, J.C., and Kim, S.J. (2007) Cloning and characterization of three epoxide hydrolases from a marine bacterium, *Erythrobacter litoralis* HTCC2594. *Appl Microbiol Biotechnol* **76**: 365–375.
- Woo, J.H., Kwon, T.H., Kim, J.T., Kim, C.G., and Lee, E.Y. (2013) Identification and characterization of epoxide hydrolase activity of polycyclic aromatic hydrocarbon-degrading bacteria for biocatalytic resolution of racemic styrene oxide and styrene oxide derivatives. *Biotechnol Lett* **35**: 599–606.
- Ye, M., Li, G., Liang, W.Q., and Liu, Y.H. (2010) Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. *Appl Microbiol Biotechnol* **87**: 1023–1031.